7 | THE CENTRAL LIMIT THEOREM

Figure 7.1 If you want to figure out the distribution of the change people carry in their pockets, using the central limit theorem and assuming your sample is large enough, you will find that the distribution is normal and bell-shaped. (credit: John Lodder)

Introduction

Chapter Objectives

By the end of this chapter, the student should be able to:

- Recognize central limit theorem problems.
- Classify continuous word problems by their distributions.
- Apply and interpret the central limit theorem for means.
- Apply and interpret the central limit theorem for sums.

Why are we so concerned with means? Two reasons are: they give us a middle ground for comparison, and they are easy to calculate. In this chapter, you will study means and the **central limit theorem**.

The **central limit theorem** (clt for short) is one of the most powerful and useful ideas in all of statistics. There are two alternative forms of the theorem, and both alternatives are concerned with drawing finite samples size *n* from a population with a known mean, μ, and a known standard deviation, σ. The first alternative says that if we collect samples of size *n* with

a "large enough *n*," calculate each sample's mean, and create a histogram of those means, then the resulting histogram will tend to have an approximate normal bell shape. The second alternative says that if we again collect samples of size *n* that are "large enough," calculate the sum of each sample and create a histogram, then the resulting histogram will again tend to have a normal bell-shape.

In either case, it does not matter what the distribution of the original population is, or whether you even need to know it. The important fact is that the distribution of sample means and the sums tend to follow the normal distribution.

The size of the sample, *n*, that is required in order to be "large enough" depends on the original population from which the samples are drawn (the sample size should be at least 30 or the data should come from a normal distribution). If the original population is far from normal, then more observations are needed for the sample means or sums to be normal. **Sampling is done with replacement.**

Collaborative Exercise

Suppose eight of you roll one fair die ten times, seven of you roll two fair dice ten times, nine of you roll five fair dice ten times, and 11 of you roll ten fair dice ten times.

Each time a person rolls more than one die, he or she calculates the sample **mean** of the faces showing. For example, one person might roll five fair dice and get 2, 2, 3, 4, 6 on one roll.

The mean is $\frac{2+2+3+4+6}{5}$ = 3.4. The 3.4 is one mean when five fair dice are rolled. This same person would roll the five dice nine more times and calculate nine more means for a total of ten means.

Your instructor will pass out the dice to several people. Roll your dice ten times. For each roll, record the faces, and find the mean. Round to the nearest 0.5.

Your instructor (and possibly you) will produce one graph (it might be a histogram) for one die, one graph for two dice, one graph for five dice, and one graph for ten dice. Since the "mean" when you roll one die is just the face on the die, what distribution do these **means** appear to be representing?

Draw the graph for the means using two dice. Do the sample means show any kind of pattern?

Draw the graph for the means using five dice. Do you see any pattern emerging?

Finally, draw the graph for the means using ten dice. Do you see any pattern to the graph? What can you conclude as you increase the number of dice?

As the number of dice rolled increases from one to two to five to ten, the following is happening:

1. The mean of the sample means remains approximately the same.
2. The spread of the sample means (the standard deviation of the sample means) gets smaller.
3. The graph appears steeper and thinner.

You have just demonstrated the central limit theorem (clt).

The central limit theorem tells you that as you increase the number of dice, **the sample means tend toward a normal distribution (the sampling distribution).**

7.1 | The Central Limit Theorem for Sample Means (Averages)

Suppose X is a random variable with a distribution that may be known or unknown (it can be any distribution). Using a subscript that matches the random variable, suppose:

a. μ_X = the mean of X
b. σ_X = the standard deviation of X

If you draw random samples of size *n*, then as *n* increases, the random variable $\bar{X}$ which consists of sample means, tends to be **normally distributed** and

$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n}}\right)$.

The **central limit theorem** for sample means says that if you keep drawing larger and larger samples (such as rolling one, two, five, and finally, ten dice) and **calculating their means,** the sample means form their own **normal distribution** (the sampling distribution). The normal distribution has the same mean as the original distribution and a variance that equals the original variance divided by, the sample size. The variable n is the number of values that are averaged together, not the number of times the experiment is done.

To put it more formally, if you draw random samples of size n, the distribution of the random variable $\bar{X}$, which consists of sample means, is called the **sampling distribution of the mean**. The sampling distribution of the mean approaches a normal distribution as n, the **sample size**, increases.

The random variable $\bar{X}$ has a different z-score associated with it from that of the random variable X. The mean $\bar{x}$ is the value of $\bar{X}$ in one sample.

$$z = \frac{\bar{x} - \mu_X}{\left(\frac{\sigma_X}{\sqrt{n}}\right)}$$

μ_X is the average of both X and $\bar{X}$.

$\sigma_{\bar{x}} = \frac{\sigma_X}{\sqrt{n}}$ = standard deviation of $\bar{X}$ and is called the **standard error of the mean.**

Using the TI-83, 83+, 84, 84+ Calculator

To find probabilities for means on the calculator, follow these steps.

2nd DISTR
2:normalcdf

$$normalcdf\left(lower\ value\ of\ the\ area,\ upper\ value\ of\ the\ area,\ mean,\ \frac{standard\ \ deviation}{\sqrt{sample\ \ size}}\right)$$

where:

- *mean* is the mean of the original distribution
- *standard deviation* is the standard deviation of the original distribution
- *sample size* = n

Example 7.1

An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size n = 25 are drawn randomly from the population.

a. Find the probability that the **sample mean** is between 85 and 92.

Solution 7.1

a. Let X = one value from the original unknown population. The probability question asks you to find a probability for the **sample mean**.

Let $\bar{X}$ = the mean of a sample of size 25. Since μ_X = 90, σ_X = 15, and n = 25,

$\bar{X} \sim N\left(90, \frac{15}{\sqrt{25}}\right)$.

Find $P(85 < \bar{x} < 92)$. Draw a graph.

$P(85 < \bar{x} < 92) = 0.6997$

The probability that the sample mean is between 85 and 92 is 0.6997.

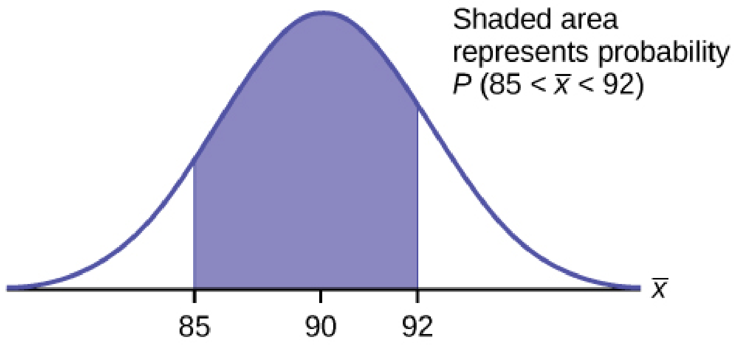


Figure 7.2

Using the TI-83, 83+, 84, 84+ Calculator

normalcdf(lower value, upper value, mean, standard error of the mean)

The parameter list is abbreviated (lower value, upper value, μ, $\frac{\sigma}{\sqrt{n}}$)

normalcdf(85,92,90, $\frac{15}{\sqrt{25}}$) = 0.6997

b. Find the value that is two standard deviations above the expected value, 90, of the sample mean.

Solution 7.1

b. To find the value that is two standard deviations above the expected value 90, use the formula:

$$\text{value} = \mu_x + (\#\text{ofTSDEVs})\left(\frac{\sigma_x}{\sqrt{n}}\right)$$

$$\text{value} = 90 + 2\left(\frac{15}{\sqrt{25}}\right) = 96$$

The value that is two standard deviations above the expected value is 96.

The standard error of the mean is $\frac{\sigma_x}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3$. Recall that the standard error of the mean is a description of how far (on average) that the sample mean will be from the population mean in repeated simple random samples of size n.

Try It Σ

7.1 An unknown distribution has a mean of 45 and a standard deviation of eight. Samples of size n = 30 are drawn randomly from the population. Find the probability that the sample mean is between 42 and 50.

Example 7.2

The length of time, in hours, it takes an "over 40" group of people to play one soccer match is normally distributed with a **mean of two hours** and a **standard deviation of 0.5 hours**. A **sample of size *n* = 50** is drawn randomly from the population. Find the probability that the **sample mean** is between 1.8 hours and 2.3 hours.

Solution 7.2

Let X = the time, in hours, it takes to play one soccer match.

The probability question asks you to find a probability for the **sample mean time, in hours**, it takes to play one soccer match.

Let $\bar{X}$ = the **mean** time, in hours, it takes to play one soccer match.

If μ_X = _________, σ_X = __________, and n = ___________, then $X \sim N$(______, ______) by the **central limit theorem for means**.

$\mu_X = 2$, $\sigma_X = 0.5$, $n = 50$, and $X \sim N\left(2, \frac{0.5}{\sqrt{50}}\right)$

Find $P(1.8 < \bar{x} < 2.3)$. Draw a graph.

$P(1.8 < \bar{x} < 2.3) = 0.9977$

$\texttt{normalcdf}\left(1.8,2.3,2,\frac{.5}{\sqrt{50}}\right) = 0.9977$

The probability that the mean time is between 1.8 hours and 2.3 hours is 0.9977.

Try It Σ

7.2 The length of time taken on the SAT for a group of students is normally distributed with a mean of 2.5 hours and a standard deviation of 0.25 hours. A sample size of n = 60 is drawn randomly from the population. Find the probability that the sample mean is between two hours and three hours.

Using the TI-83, 83+, 84, 84+ Calculator

To find percentiles for means on the calculator, follow these steps.

2nd DIStR
3:invNorm

$k = \text{invNorm}\left(\text{area to the left of } k, \text{mean}, \frac{standard\ \ deviation}{\sqrt{sample\ \ size}}\right)$

where:

- k = the k^{th} percentile
- *mean* is the mean of the original distribution
- *standard deviation* is the standard deviation of the original distribution
- *sample size* = n

Example 7.3

In a recent study reported Oct. 29, 2012 on the Flurry Blog, the mean age of tablet users is 34 years. Suppose the standard deviation is 15 years. Take a sample of size $n = 100$.

a. What are the mean and standard deviation for the sample mean ages of tablet users?

b. What does the distribution look like?

c. Find the probability that the sample mean age is more than 30 years (the reported mean age of tablet users in this particular study).

d. Find the 95$^{\text{th}}$ percentile for the sample mean age (to one decimal place).

Solution 7.3

a. Since the sample mean tends to target the population mean, we have $\mu_X = \mu = 34$. The sample standard deviation is given by $\sigma_X = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$

b. The central limit theorem states that for large sample sizes(n), the sampling distribution will be approximately normal.

c. The probability that the sample mean age is more than 30 is given by $P(X > 30)$ = `normalcdf`(30,E99,34,1.5) = 0.9962

d. Let k = the 95$^{\text{th}}$ percentile.
$k = \text{invNorm}\left(0.95, 34, \frac{15}{\sqrt{100}}\right) = 36.5$

Try It Σ

7.3 In an article on Flurry Blog, a gaming marketing gap for men between the ages of 30 and 40 is identified. You are researching a startup game targeted at the 35-year-old demographic. Your idea is to develop a strategy game that can be played by men from their late 20s through their late 30s. Based on the article's data, industry research shows that the average strategy player is 28 years old with a standard deviation of 4.8 years. You take a sample of 100 randomly selected gamers. If your target market is 29- to 35-year-olds, should you continue with your development strategy?

Example 7.4

The mean number of minutes for app engagement by a tablet user is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample of 60.

a. What are the mean and standard deviation for the sample mean number of app engagement by a tablet user?

b. What is the standard error of the mean?

c. Find the 90$^{\text{th}}$ percentile for the sample mean time for app engagement for a tablet user. Interpret this value in a complete sentence.

d. Find the probability that the sample mean is between eight minutes and 8.5 minutes.

Solution 7.4

a. $\mu_{\bar{x}} = \mu = 8.2 \ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{60}} = 0.13$

b. This allows us to calculate the probability of sample means of a particular distance from the mean, in repeated samples of size 60.

c. Let k = the 90th percentile

k = `invNorm` $\left(0.90,8.2,\frac{1}{\sqrt{60}}\right)$ = 8.37. This values indicates that 90 percent of the average app engagement time for table users is less than 8.37 minutes.

d. $P(8 < \bar{x} < 8.5)$ = `normalcdf` $\left(8,8.5,8.2,\frac{1}{\sqrt{60}}\right)$ = 0.9293

Try It Σ

7.4 Cans of a cola beverage claim to contain 16 ounces. The amounts in a sample are measured and the statistics are $n = 34$, $\bar{x} = 16.01$ ounces. If the cans are filled so that $\mu = 16.00$ ounces (as labeled) and $\sigma = 0.143$ ounces, find the probability that a sample of 34 cans will have an average amount greater than 16.01 ounces. Do the results suggest that cans are filled with an amount greater than 16 ounces?

7.2 | The Central Limit Theorem for Sums

Suppose X is a random variable with a distribution that may be **known or unknown** (it can be any distribution) and suppose:

a. μ_X = the mean of X

b. σ_X = the standard deviation of X

If you draw random samples of size n, then as n increases, the random variable ΣX consisting of sums tends to be **normally distributed** and $\Sigma X \sim N((n)(\mu_X), (\sqrt{n})(\sigma_X))$.

The central limit theorem for sums says that if you keep drawing larger and larger samples and taking their sums, the sums form their own normal distribution (the sampling distribution), which approaches a normal distribution as the sample size increases. **The normal distribution has a mean equal to the original mean multiplied by the sample size and a standard deviation equal to the original standard deviation multiplied by the square root of the sample size.**

The random variable ΣX has the following z-score associated with it:

a. Σx is one sum.

b. $z = \dfrac{\Sigma x - (n)(\mu_X)}{(\sqrt{n})(\sigma_X)}$

 i. $(n)(\mu_X)$ = the mean of ΣX

 ii. $(\sqrt{n})(\sigma_X)$ = standard deviation of ΣX

Using the TI-83, 83+, 84, 84+ Calculator

To find probabilities for sums on the calculator, follow these steps.

2nd `DISTR`
2:`normalcdf`
`normalcdf`(lower value of the area, upper value of the area, (n)(mean), $(\sqrt{n})$(standard deviation))

where:

- *mean* is the mean of the original distribution
- *standard deviation* is the standard deviation of the original distribution
- ***sample size*** = n

Example 7.5

An unknown distribution has a mean of 90 and a standard deviation of 15. A sample of size 80 is drawn randomly from the population.

a. Find the probability that the sum of the 80 values (or the total of the 80 values) is more than 7,500.

b. Find the sum that is 1.5 standard deviations above the mean of the sums.

Solution 7.5

Let X = one value from the original unknown population. The probability question asks you to find a probability for **the sum (or total of) 80 values.**

ΣX = the sum or total of 80 values. Since $\mu_X = 90$, $\sigma_X = 15$, and $n = 80$, $\Sigma X \sim N((80)(90), (\sqrt{80})(15))$

- mean of the sums = $(n)(\mu_X) = (80)(90) = 7{,}200$
- standard deviation of the sums = $(\sqrt{n})(\sigma_X) = (\sqrt{80})(15)$
- sum of 80 values = $\Sigma x = 7{,}500$

a. Find $P(\Sigma x > 7{,}500)$

$P(\Sigma x > 7{,}500) = 0.0127$

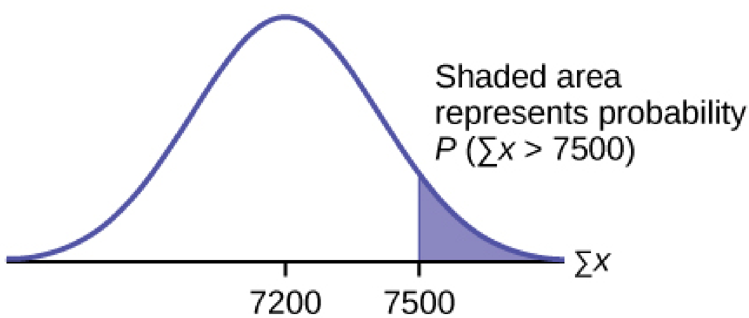


Figure 7.3

Using the TI-83, 83+, 84, 84+ Calculator

`normalcdf`(lower value, upper value, mean of sums, `stdev` of sums)

The parameter list is abbreviated(lower, upper, $(n)(\mu_X$, $(\sqrt{n})(\sigma_X)$)

`normalcdf` (7500,1E99,(80)(90),$(\sqrt{80})$(15)) = 0.0127

REMINDER

1E99 = 10^{99}.

Press the `EE` key for E.

b. Find Σx where $z = 1.5$.

$\Sigma x = (n)(\mu_X) + (z)(\sqrt{n})(\sigma_X) = (80)(90) + (1.5)(\sqrt{80})(15) = 7{,}401.2$

Try It Σ

7.5 An unknown distribution has a mean of 45 and a standard deviation of eight. A sample size of 50 is drawn randomly from the population. Find the probability that the sum of the 50 values is more than 2,400.

Using the TI-83, 83+, 84, 84+ Calculator

To find percentiles for sums on the calculator, follow these steps.

2nd DIStR
3:invNorm
k = invNorm (area to the left of k, (n)(mean), ($\sqrt{n}$) (standard deviation))

where:

- k is the k^{th} **percentile**
- *mean* is the mean of the original distribution
- *standard deviation* is the standard deviation of the original distribution
- *sample size* = n

Example 7.6

In a recent study reported Oct. 29, 2012 on the Flurry Blog, the mean age of tablet users is 34 years. Suppose the standard deviation is 15 years. The sample of size is 50.

a. What are the mean and standard deviation for the sum of the ages of tablet users? What is the distribution?

b. Find the probability that the sum of the ages is between 1,500 and 1,800 years.

c. Find the 80th percentile for the sum of the 50 ages.

Solution 7.6

a. $\mu_{\Sigma x} = n\mu_x = 50(34) = 1{,}700$ and $\sigma_{\Sigma x} = \sqrt{n}\,\sigma_x = (\sqrt{50})(15) = 106.01$

The distribution is normal for sums by the central limit theorem.

b. $P(1500 < \Sigma x < 1800) =$ `normalcdf` (1,500, 1,800, (50)(34), $(\sqrt{50})(15)$) = 0.7974

c. Let k = the 80th percentile.
$k =$ `invNorm`(0.80,(50)(34), $(\sqrt{50})(15)$) = 1,789.3

Try It Σ

7.6 In a recent study reported Oct.29, 2012 on the Flurry Blog, the mean age of tablet users is 35 years. Suppose the standard deviation is ten years. The sample size is 39.

a. What are the mean and standard deviation for the sum of the ages of tablet users? What is the distribution?

b. Find the probability that the sum of the ages is between 1,400 and 1,500 years.

c. Find the 90th percentile for the sum of the 39 ages.

Example 7.7

The mean number of minutes for app engagement by a tablet user is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample of size 70.

a. What are the mean and standard deviation for the sums?

b. Find the 95th percentile for the sum of the sample. Interpret this value in a complete sentence.

c. Find the probability that the sum of the sample is at least ten hours.

Solution 7.7

a. $\mu_{\Sigma x} = n\mu_x = 70(8.2) = 574$ minutes and $\sigma_{\Sigma x} = (\sqrt{n})(\sigma_x) = (\sqrt{70})(1) = 8.37$ minutes

b. Let k = the 95th percentile.
$k =$ invNorm (0.95,(70)(8.2), $(\sqrt{70})(1)$) = 587.76 minutes

Ninety five percent of the app engagement times are at most 587.76 minutes.

c. ten hours = 600 minutes
$P(\Sigma x \ge 600) =$ `normalcdf`(600,E99,(70)(8.2), $(\sqrt{70})(1)$) = 0.0009

Try It Σ

7.7 The mean number of minutes for app engagement by a table use is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample size of 70.

a. What is the probability that the sum of the sample is between seven hours and ten hours? What does this mean in context of the problem?

b. Find the 84^{th} and 16^{th} percentiles for the sum of the sample. Interpret these values in context.

7.3 | Using the Central Limit Theorem

It is important for you to understand when to use the **central limit theorem**. If you are being asked to find the probability of the mean, use the clt for the mean. If you are being asked to find the probability of a sum or total, use the clt for sums. This also applies to percentiles for means and sums.

NOTE

If you are being asked to find the probability of an **individual** value, do **not** use the clt. **Use the distribution of its random variable.**

Examples of the Central Limit Theorem

Law of Large Numbers

The **law of large numbers** says that if you take samples of larger and larger size from any population, then the mean $\bar{x}$ of the sample tends to get closer and closer to μ. From the central limit theorem, we know that as n gets larger and larger, the sample means follow a normal distribution. The larger n gets, the smaller the standard deviation gets. (Remember that the standard deviation for $\bar{X}$ is $\frac{\sigma}{\sqrt{n}}$.) This means that the sample mean $\bar{x}$ must be close to the population mean μ. We can say that μ is the value that the sample means approach as n gets larger. The central limit theorem illustrates the law of large numbers.

Central Limit Theorem for the Mean and Sum Examples

Example 7.8

A study involving stress is conducted among the students on a college campus. **The stress scores follow a uniform distribution** with the lowest stress score equal to one and the highest equal to five. Using a sample of 75 students, find:

a. The probability that the **mean stress score** for the 75 students is less than two.

b. The 90^{th} percentile for the **mean stress score** for the 75 students.

c. The probability that the **total of the 75 stress scores** is less than 200.

d. The 90^{th} percentile for the **total stress score** for the 75 students.

Let X = one stress score.

Problems a and b ask you to find a probability or a percentile for a **mean**. Problems c and d ask you to find a probability or a percentile for a **total or sum**. The sample size, n, is equal to 75.

Since the individual stress scores follow a uniform distribution, $X \sim U(1, 5)$ where $a = 1$ and $b = 5$ (See **Continuous Random Variables** for an explanation on the uniform distribution).

$$\mu_X = \frac{a+b}{2} = \frac{1+5}{2} = 3$$

$$\sigma_X = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(5-1)^2}{12}} = 1.15$$

For problems 1. and 2., let $\bar{X}$ = the mean stress score for the 75 students. Then,

$\bar{X} \sim N\left(3, \frac{1.15}{\sqrt{75}}\right)$ where $n = 75$.

a. Find $P(\bar{x} < 2)$. Draw the graph.

Solution 7.8

a. $P(\bar{x} < 2) = 0$

The probability that the mean stress score is less than two is about zero.

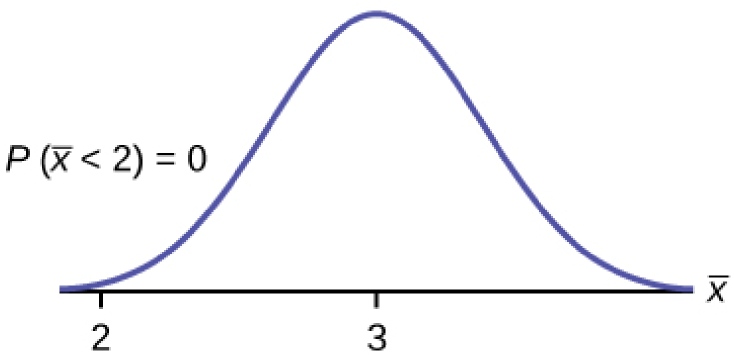


Figure 7.4

normalcdf $\left(1,2,3,\frac{1.15}{\sqrt{75}}\right) = 0$

REMINDER

The smallest stress score is one.

b. Find the 90th percentile for the mean of 75 stress scores. Draw a graph.

Solution 7.8

b. Let k = the 90th precentile.

Find k, where $P(\bar{x} < k) = 0.90$.

$k = 3.2$

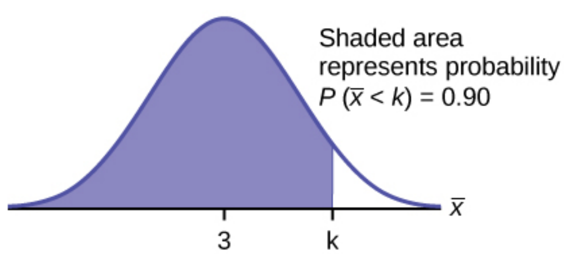


Figure 7.5

The 90th percentile for the mean of 75 scores is about 3.2. This tells us that 90% of all the means of 75 stress scores are at most 3.2, and that 10% are at least 3.2.

`invNorm` $\left(0.90,3,\frac{1.15}{\sqrt{75}}\right)$ = 3.2

For problems c and d, let ΣX = the sum of the 75 stress scores. Then, $\Sigma X \sim N[(75)(3), (\sqrt{75})\ (1.15)]$

c. Find $P(\Sigma x < 200)$. Draw the graph.

Solution 7.8

c. The mean of the sum of 75 stress scores is (75)(3) = 225

The standard deviation of the sum of 75 stress scores is $(\sqrt{75})\ (1.15) = 9.96$

$P(\Sigma x < 200) = 0$

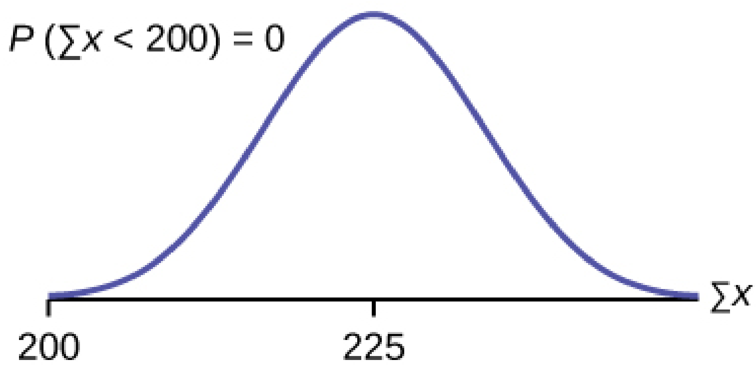


Figure 7.6

The probability that the total of 75 scores is less than 200 is about zero.

`normalcdf` (75,200,(75)(3), $(\sqrt{75})$ (1.15)).

REMINDER

The smallest total of 75 stress scores is 75, because the smallest single score is one.

d. Find the 90th percentile for the total of 75 stress scores. Draw a graph.

Solution 7.8

d. Let k = the 90th percentile.

Find k where $P(\Sigma x < k) = 0.90$.

$k = 237.8$

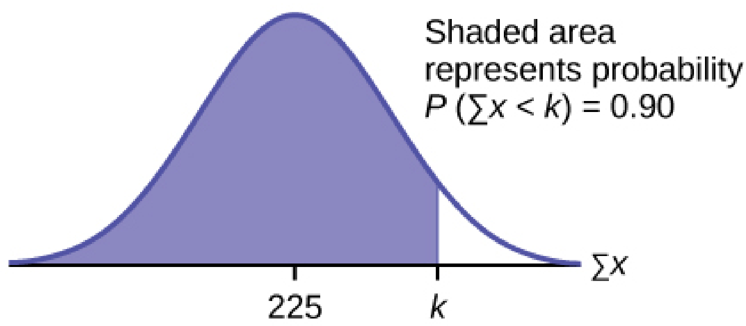


Figure 7.7

The 90th percentile for the sum of 75 scores is about 237.8. This tells us that 90% of all the sums of 75 scores are no more than 237.8 and 10% are no less than 237.8.

`invNorm`(0.90,(75)(3), ($\sqrt{75}$) (1.15)) = 237.8

Try It Σ

7.8 Use the information in **Example 7.8**, but use a sample size of 55 to answer the following questions.

a. Find $P(\bar{x} < 7)$.

b. Find $P(\Sigma x > 170)$.

c. Find the 80th percentile for the mean of 55 scores.

d. Find the 85th percentile for the sum of 55 scores.

Example 7.9

Suppose that a market research analyst for a cell phone company conducts a study of their customers who exceed the time allowance included on their basic cell phone contract; the analyst finds that for those people who exceed the time included in their basic contract, the **excess time used** follows an **exponential distribution** with a mean of 22 minutes.

Consider a random sample of 80 customers who exceed the time allowance included in their basic cell phone contract.

Let X = the excess time used by one INDIVIDUAL cell phone customer who exceeds his contracted time allowance.

$X \sim Exp\left(\frac{1}{22}\right)$. From previous chapters, we know that $\mu = 22$ and $\sigma = 22$.

Let $\bar{X}$ = the mean excess time used by a sample of $n = 80$ customers who exceed their contracted time allowance.

$\bar{X} \sim N\left(22, \frac{22}{\sqrt{80}}\right)$ by the central limit theorem for sample means

Using the clt to find probability

a. Find the probability that the mean excess time used by the 80 customers in the sample is longer than 20 minutes. This is asking us to find $P(\bar{x} > 20)$. Draw the graph.

b. Suppose that one customer who exceeds the time limit for his cell phone contract is randomly selected. Find the probability that this individual customer's excess time is longer than 20 minutes. This is asking us to find $P(x > 20)$.

c. Explain why the probabilities in parts a and b are different.

Solution 7.9

a. Find: $P(\bar{x} > 20)$

$P(\bar{x} > 20) = 0.79199$ using `normalcdf` $\left(20,1E99,22,\frac{22}{\sqrt{80}}\right)$

The probability is 0.7919 that the mean excess time used is more than 20 minutes, for a sample of 80 customers who exceed their contracted time allowance.

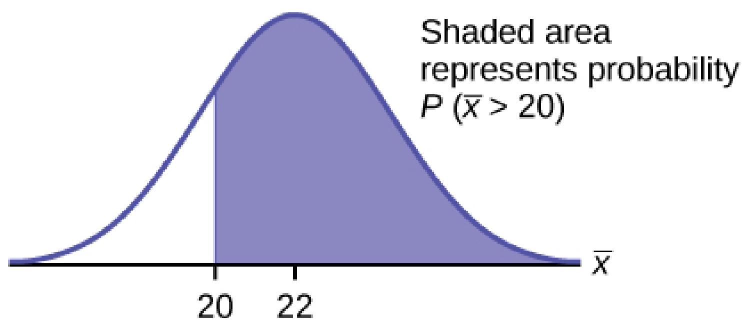


Figure 7.8

REMINDER

1E99 = 10^{99} and –1E99 = –10^{99}. Press the `EE` key for E. Or just use 10^{99} instead of 1E99.

b. Find $P(x > 20)$. Remember to use the exponential distribution for an **individual:** $X \sim Exp\left(\frac{1}{22}\right)$.

$$P(x > 20) = e^{\left(-\left(\frac{1}{22}\right)(20)\right)} \text{ or } e^{(-0.04545(20))} = 0.4029$$

c. 1. $P(x > 20) = 0.4029$ but $P(\bar{x} > 20) = 0.7919$

2. The probabilities are not equal because we use different distributions to calculate the probability for individuals and for means.

3. When asked to find the probability of an individual value, use the stated distribution of its random variable; do not use the clt. Use the clt with the normal distribution when you are being asked to find the probability for a mean.

Using the clt to find percentiles Find the 95th percentile for the **sample mean excess time** for samples of 80 customers who exceed their basic contract time allowances. Draw a graph.

Solution 7.9

Let k = the 95th percentile. Find k where $P(\bar{x} < k) = 0.95$

$k = 26.0$ using `invNorm` $\left(0.95,22,\frac{22}{\sqrt{80}}\right) = 26.0$

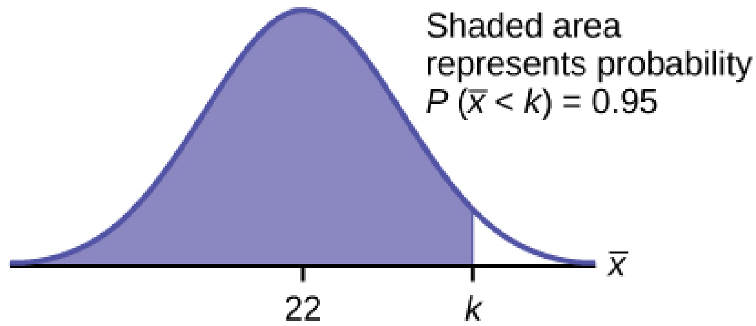


Figure 7.9

The 95th percentile for the **sample mean excess time used** is about 26.0 minutes for random samples of 80 customers who exceed their contractual allowed time.

Ninety five percent of such samples would have means under 26 minutes; only five percent of such samples would have means above 26 minutes.

Try It Σ

7.9 Use the information in **Example 7.9**, but change the sample size to 144.

a. Find $P(20 < \bar{x} < 30)$.

b. Find $P(\Sigma x$ is at least 3,000).

c. Find the 75th percentile for the sample mean excess time of 144 customers.

d. Find the 85th percentile for the sum of 144 excess times used by customers.

Example 7.10

In the United States, someone is sexually assaulted every two minutes, on average, according to a number of studies. Suppose the standard deviation is 0.5 minutes and the sample size is 100.

a. Find the median, the first quartile, and the third quartile for the sample mean time of sexual assaults in the United States.

b. Find the median, the first quartile, and the third quartile for the sum of sample times of sexual assaults in the United States.

c. Find the probability that a sexual assault occurs on the average between 1.75 and 1.85 minutes.

d. Find the value that is two standard deviations above the sample mean.

e. Find the *IQR* for the sum of the sample times.

Solution 7.10

a. We have, $\mu_x = \mu = 2$ and $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{0.5}{10} = 0.05$. Therefore:

1. 50th percentile = $\mu_x = \mu = 2$
2. 25th percentile = *invNorm*(0.25,2,0.05) = 1.97
3. 75th percentile = *invNorm*(0.75,2,0.05) = 2.03

b. We have $\mu_{\Sigma x} = n(\mu_x) = 100(2) = 200$ and $\sigma_{\mu x} = \sqrt{n}(\sigma_x) = 10(0.5) = 5$. Therefore

1. 50th percentile = $\mu_{\Sigma x} = n(\mu_x) = 100(2) = 200$
2. 25th percentile = invNorm(0.25,200,5) = 196.63
3. 75th percentile = invNorm(0.75,200,5) = 203.37

c. $P(1.75 < \bar{x} < 1.85)$ = `normalcdf`(1.75,1.85,2,0.05) = 0.0013

d. Using the z-score equation, $z = \dfrac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$, and solving for x, we have $x = 2(0.05) + 2 = 2.1$

e. The IQR is 75th percentile – 25th percentile = 203.37 – 196.63 = 6.74

Try It Σ

7.10 Based on data from the National Health Survey, women between the ages of 18 and 24 have an average systolic blood pressures (in mm Hg) of 114.8 with a standard deviation of 13.1. Systolic blood pressure for women between the ages of 18 to 24 follow a normal distribution.

a. If one woman from this population is randomly selected, find the probability that her systolic blood pressure is greater than 120.

b. If 40 women from this population are randomly selected, find the probability that their mean systolic blood pressure is greater than 120.

c. If the sample were four women between the ages of 18 to 24 and we did not know the original distribution, could the central limit theorem be used?

Example 7.11

A study was done about violence against prostitutes and the symptoms of the posttraumatic stress that they developed. The age range of the prostitutes was 14 to 61. The mean age was 30.9 years with a standard deviation of nine years.

a. In a sample of 25 prostitutes, what is the probability that the mean age of the prostitutes is less than 35?

b. Is it likely that the mean age of the sample group could be more than 50 years? Interpret the results.

c. In a sample of 49 prostitutes, what is the probability that the sum of the ages is no less than 1,600?

d. Is it likely that the sum of the ages of the 49 prostitutes is at most 1,595? Interpret the results.

e. Find the 95th percentile for the sample mean age of 65 prostitutes. Interpret the results.

f. Find the 90th percentile for the sum of the ages of 65 prostitutes. Interpret the results.

Solution 7.11

a. $P(\bar{x} < 35)$ = `normalcdf`(-E99,35,30.9,1.8) = 0.9886

b. $P(\bar{x} > 50)$ = `normalcdf`(50, E99,30.9,1.8) ≈ 0. For this sample group, it is almost impossible for the group’s average age to be more than 50. However, it is still possible for an individual in this group to have an age greater than 50.

c. $P(\Sigma x \geq 1{,}600)$ = `normalcdf`(1600,E99,1514.10,63) = 0.0864

d. $P(\Sigma x \leq 1{,}595)$ = `normalcdf`(-E99,1595,1514.10,63) = 0.9005. This means that there is a 90% chance that the sum of the ages for the sample group n = 49 is at most 1595.

e. The 95th percentile = `invNorm`(0.95,30.9,1.1) = 32.7. This indicates that 95% of the prostitutes in the sample of 65 are younger than 32.7 years, on average.

f. The 90th percentile = `invNorm`(0.90,2008.5,72.56) = 2101.5. This indicates that 90% of the prostitutes in the sample of 65 have a sum of ages less than 2,101.5 years.

Try It Σ

7.11 According to Boeing data, the 757 airliner carries 200 passengers and has doors with a mean height of 72 inches. Assume for a certain population of men we have a mean of 69.0 inches and a standard deviation of 2.8 inches.

a. What mean doorway height would allow 95% of men to enter the aircraft without bending?

b. Assume that half of the 200 passengers are men. What mean doorway height satisfies the condition that there is a 0.95 probability that this height is greater than the mean height of 100 men?

c. For engineers designing the 757, which result is more relevant: the height from part a or part b? Why?

HISTORICAL NOTE

: Normal Approximation to the Binomial

Historically, being able to compute binomial probabilities was one of the most important applications of the central limit theorem. Binomial probabilities with a small value for n(say, 20) were displayed in a table in a book. To calculate the probabilities with large values of n, you had to use the binomial formula, which could be very complicated. Using the **normal approximation to the binomial** distribution simplified the process. To compute the normal approximation to the binomial distribution, take a simple random sample from a population. You must meet the conditions for a **binomial distribution**:

- there are a certain number n of independent trials
- the outcomes of any trial are success or failure
- each trial has the same probability of a success p

Recall that if X is the binomial random variable, then $X \sim B(n, p)$. The shape of the binomial distribution needs to be similar to the shape of the normal distribution. To ensure this, the quantities np and nq must both be greater than five ($np > 5$ and $nq > 5$; the approximation is better if they are both greater than or equal to 10). Then the binomial can be approximated by the normal distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$. Remember that $q = 1 – p$. In order to get the best approximation, add 0.5 to x or subtract 0.5 from x (use $x + 0.5$ or $x – 0.5$). The number 0.5 is called the **continuity correction factor** and is used in the following example.

Example 7.12

Suppose in a local Kindergarten through 12th grade (K - 12) school district, 53 percent of the population favor a charter school for grades K through 5. A simple random sample of 300 is surveyed.

a. Find the probability that **at least 150** favor a charter school.

b. Find the probability that **at most 160** favor a charter school.

c. Find the probability that **more than 155** favor a charter school.

d. Find the probability that **fewer than 147** favor a charter school.

e. Find the probability that **exactly 175** favor a charter school.

Let X = the number that favor a charter school for grades K trough 5. $X \sim B(n, p)$ where n = 300 and p = 0.53. Since $np > 5$ and $nq > 5$, use the normal approximation to the binomial. The formulas for the mean and standard deviation are $\mu = np$ and $\sigma = \sqrt{npq}$. The mean is 159 and the standard deviation is 8.6447. The random variable for the normal distribution is Y. $Y \sim N(159, 8.6447)$. See **The Normal Distribution** for help with calculator instructions.

For part a, you **include 150** so $P(X \geq 150)$ has normal approximation $P(Y \geq 149.5) = 0.8641$.

`normalcdf`(149.5,10^99,159,8.6447) = 0.8641.

For part b, you **include 160** so $P(X \leq 160)$ has normal appraximation $P(Y \leq 160.5) = 0.5689$.

`normalcdf`(0,160.5,159,8.6447) = 0.5689

For part c, you **exclude 155** so $P(X > 155)$ has normal approximation $P(y > 155.5) = 0.6572$.

`normalcdf`(155.5,10^99,159,8.6447) = 0.6572.

For part d, you **exclude 147** so $P(X < 147)$ has normal approximation $P(Y < 146.5) = 0.0741$.

`normalcdf`(0,146.5,159,8.6447) = 0.0741

For part e,$P(X = 175)$ has normal approximation $P(174.5 < Y < 175.5) = 0.0083$.

`normalcdf`(174.5,175.5,159,8.6447) = 0.0083

Because of calculators and computer software that let you calculate binomial probabilities for large values of *n* easily, it is not necessary to use the the normal approximation to the binomial distribution, provided that you have access to these technology tools. Most school labs have Microsoft Excel, an example of computer software that calculates binomial probabilities. Many students have access to the TI-83 or 84 series calculators, and they easily calculate probabilities for the binomial distribution. If you type in "binomial probability distribution calculation" in an Internet browser, you can find at least one online calculator for the binomial.

For **Example 7.10**, the probabilities are calculated using the following binomial distribution: ($n = 300$ and $p = 0.53$). Compare the binomial and normal distribution answers. See **Discrete Random Variables** for help with calculator instructions for the binomial.

$P(X \geq 150)$:`1 - binomialcdf`(300,0.53,149) = 0.8641

$P(X \leq 160)$:`binomialcdf`(300,0.53,160) = 0.5684

$P(X > 155)$:`1 - binomialcdf`(300,0.53,155) = 0.6576

$P(X < 147)$:`binomialcdf`(300,0.53,146) = 0.0742

$P(X = 175)$:(You use the binomial pdf.)`binomialpdf`(300,0.53,175) = 0.0083

Try It Σ

7.12 In a city, 46 percent of the population favor the incumbent, Dawn Morgan, for mayor. A simple random sample of 500 is taken. Using the continuity correction factor, find the probability that at least 250 favor Dawn Morgan for mayor.

7.4 | Central Limit Theorem (Pocket Change)

Stats Lab

7.1 Central Limit Theorem (Pocket Change)

Class Time:

Names:

Student Learning Outcomes

- The student will demonstrate and compare properties of the central limit theorem.

NOTE

This lab works best when sampling from several classes and combining data.

Collect the Data

1. Count the change in your pocket. (Do not include bills.)
2. Randomly survey 30 classmates. Record the values of the change in **Table 7.1**.

________	________	________	________	________
________	________	________	________	________
________	________	________	________	________
________	________	________	________	________
________	________	________	________	________
________	________	________	________	________

Table 7.1

3. Construct a histogram. Make five to six intervals. Sketch the graph using a ruler and pencil. Scale the axes.

Figure 7.10

4. Calculate the following (n = 1; surveying one person at a time):
   a. $\bar{x}$ = _______
   b. s = _______
5. Draw a smooth curve through the tops of the bars of the histogram. Use one to two complete sentences to describe the general shape of the curve.

Collecting Averages of Pairs

Repeat steps one through five of the section **Collect the Data.** with one exception. Instead of recording the change of 30 classmates, record the average change of 30 pairs.

1. Randomly survey 30 **pairs** of classmates.
2. Record the values of the average of their change in **Table 7.2**.

Table 7.2

3. Construct a histogram. Scale the axes using the same scaling you used for the section titled **Collect the Data**. Sketch the graph using a ruler and a pencil.

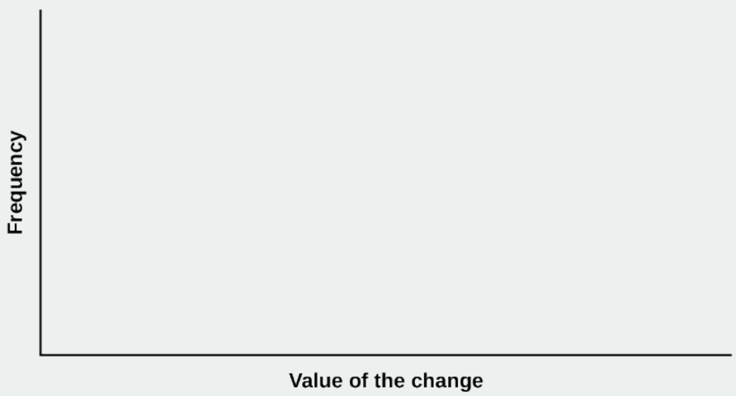


Figure 7.11

4. Calculate the following (n = 2; surveying two people at a time):
   a. $\bar{x}$ = _______
   b. s = _______
5. Draw a smooth curve through tops of the bars of the histogram. Use one to two complete sentences to describe the general shape of the curve.

Collecting Averages of Groups of Five

Repeat steps one through five (of the section titled **Collect the Data**) with one exception. Instead of recording the change of 30 classmates, record the average change of 30 groups of five.

1. Randomly survey 30 **groups of five** classmates.
2. Record the values of the average of their change.

________	________	________	________	________
________	________	________	________	________
________	________	________	________	________
________	________	________	________	________

Table 7.3

3. Construct a histogram. Scale the axes using the same scaling you used for the section titled **Collect the Data**. Sketch the graph using a ruler and a pencil.

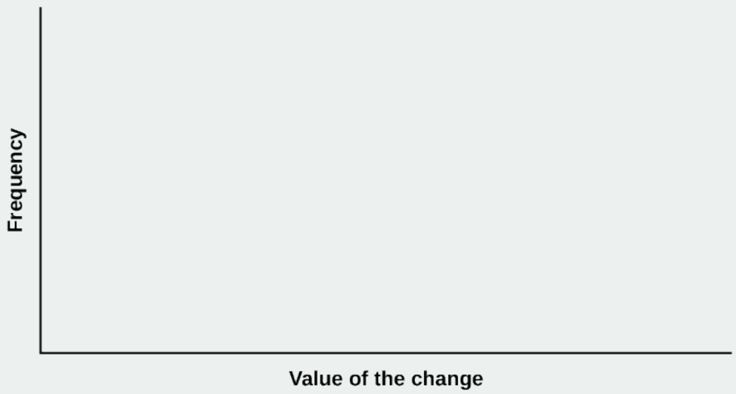


Figure 7.12

4. Calculate the following (n = 5; surveying five people at a time):
   a. $\bar{x}$ = _______
   b. s = _______
5. Draw a smooth curve through tops of the bars of the histogram. Use one to two complete sentences to describe the general shape of the curve.

Discussion Questions

1. Why did the shape of the distribution of the data change, as n changed? Use one to two complete sentences to explain what happened.
2. In the section titled **Collect the Data**, what was the approximate distribution of the data? X ~ _____(_____,_____)
3. In the section titled **Collecting Averages of Groups of Five**, what was the approximate distribution of the averages? $\bar{X}$ ~ _____(_____,_____)
4. In one to two complete sentences, explain any differences in your answers to the previous two questions.

7.5 | Central Limit Theorem (Cookie Recipes)

Stats Lab

7.2 Central Limit Theorem (Cookie Recipes)

Class Time:

Names:

Student Learning Outcomes

- The student will demonstrate and compare properties of the central limit theorem.

Given

X = length of time (in days) that a cookie recipe lasted at the Olmstead Homestead. (Assume that each of the different recipes makes the same quantity of cookies.)

Recipe #	*X*		Recipe #	*X*		Recipe #	*X*		Recipe #	*X*
1	1		16	2		31	3		46	2
2	5		17	2		32	4		47	2
3	2		18	4		33	5		48	11
4	5		19	6		34	6		49	5
5	6		20	1		35	6		50	5
6	1		21	6		36	1		51	4
7	2		22	5		37	1		52	6
8	6		23	2		38	2		53	5
9	5		24	5		39	1		54	1
10	2		25	1		40	6		55	1
11	5		26	6		41	1		56	2
12	1		27	4		42	6		57	4
13	1		28	1		43	2		58	3
14	3		29	6		44	6		59	6
15	2		30	2		45	2		60	5

Table 7.4

Calculate the following:

a. μ_X = ________

b. σ_X = ________

Collect the Data

Use a random number generator to randomly select four samples of size $n = 5$ from the given population. Record your samples in **Table 7.5**. Then, for each sample, calculate the mean to the nearest tenth. Record them in the spaces provided. Record the sample means for the rest of the class.

1. Complete the table:

	Sample 1	Sample 2	Sample 3	Sample 4	Sample means from other groups:
Means:	$\bar{x}$ = ____	$\bar{x}$ = ____	$\bar{x}$ = ____	$\bar{x}$ = ____	

Table 7.5

2. Calculate the following:
   a. $\bar{x}$ = _______
   b. $s_{\bar{x}}$ = _______
3. Again, use a random number generator to randomly select four samples from the population. This time, make the samples of size n = 10. Record the samples in **Table 7.6**. As before, for each sample, calculate the mean to the nearest tenth. Record them in the spaces provided. Record the sample means for the rest of the class.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample means from other groups
Means:	$\bar{x}$ = ____	$\bar{x}$ = ____	$\bar{x}$ = ____	$\bar{x}$ = ____	

Table 7.6

4. Calculate the following:
   a. $\bar{x}$ = ______
   b. $s_{\bar{x}}$ = ______
5. For the original population, construct a histogram. Make intervals with a bar width of one day. Sketch the graph using a ruler and pencil. Scale the axes.

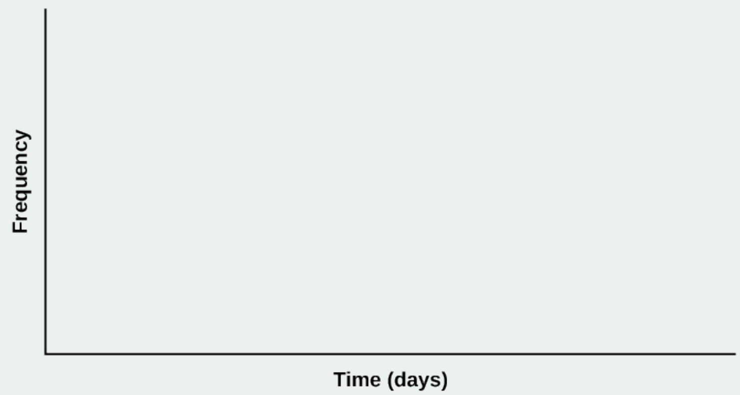


Figure 7.13

6. Draw a smooth curve through the tops of the bars of the histogram. Use one to two complete sentences to describe the general shape of the curve.

Repeat the Procedure for *n* = 5

1. For the sample of $n = 5$ days averaged together, construct a histogram of the averages (your means together with the means of the other groups). Make intervals with bar widths of $\frac{1}{2}$ a day. Sketch the graph using a ruler and pencil. Scale the axes.

Figure 7.14

2. Draw a smooth curve through the tops of the bars of the histogram. Use one to two complete sentences to describe the general shape of the curve.

Repeat the Procedure for *n* = 10

1. For the sample of $n = 10$ days averaged together, construct a histogram of the averages (your means together with the means of the other groups). Make intervals with bar widths of $\frac{1}{2}$ a day. Sketch the graph using a ruler and pencil. Scale the axes.

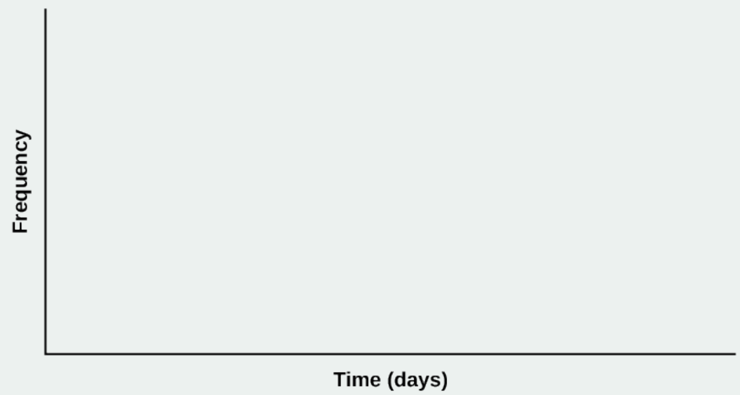


Figure 7.15

2. Draw a smooth curve through the tops of the bars of the histogram. Use one to two complete sentences to describe the general shape of the curve.

Discussion Questions

1. Compare the three histograms you have made, the one for the population and the two for the sample means. In three to five sentences, describe the similarities and differences.
2. State the theoretical (according to the clt) distributions for the sample means.
   a. $n = 5$: $\bar{x}$ ~ _____(_____,_____)
   b. $n = 10$: $\bar{x}$ ~ _____(_____,_____)
3. Are the sample means for $n = 5$ and $n = 10$ “close” to the theoretical mean, μ_x? Explain why or why not.
4. Which of the two distributions of sample means has the smaller standard deviation? Why?
5. As n changed, why did the shape of the distribution of the data change? Use one to two complete sentences to explain what happened.

KEY TERMS

Average a number that describes the central tendency of the data; there are a number of specialized averages, including the arithmetic mean, weighted mean, median, mode, and geometric mean.

Central Limit Theorem Given a random variable (RV) with known mean μ and known standard deviation, σ, we are sampling with size n, and we are interested in two new RVs: the sample mean, $\bar{X}$, and the sample sum, ΣX. If the size (n) of the sample is sufficiently large, then $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ and $\Sigma X \sim N(n\mu, (\sqrt{n})(\sigma))$. If the size ($n$) of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distributions regardless of the shape of the population. The mean of the sample means will equal the population mean, and the mean of the sample sums will equal n times the population mean. The standard deviation of the distribution of the sample means, $\frac{\sigma}{\sqrt{n}}$, is called the standard error of the mean.

Exponential Distribution a continuous random variable (RV) that appears when we are interested in the intervals of time between some random events, for example, the length of time between emergency arrivals at a hospital, notation: $X \sim Exp(m)$. The mean is $\mu = \frac{1}{m}$ and the standard deviation is $\sigma = \frac{1}{m}$. The probability density function is $f(x) = me^{-mx}$, $x \geq 0$ and the cumulative distribution function is $P(X \leq x) = 1 - e^{-mx}$.

Mean a number that measures the central tendency; a common name for mean is "average." The term "mean" is a shortened form of "arithmetic mean." By definition, the mean for a sample (denoted by $\bar{x}$) is $\bar{x} = \frac{\text{Sum of all values in the sample}}{\text{Number of values in the sample}}$, and the mean for a population (denoted by μ) is $\mu = \frac{\text{Sum of all values in the population}}{\text{Number of values in the population}}$.

Normal Distribution a continuous random variable (RV) with pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, where μ is the mean of the distribution and σ is the standard deviation; notation: $X \sim N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$, the RV is called a **standard normal distribution**.

Normal Distribution a continuous random variable (RV) with pdf $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, where μ is the mean of the distribution and σ is the standard deviation.; notation: $X \sim N(\mu, \sigma)$. If $\mu = 0$ and $\sigma = 1$, the RV is called the **standard normal distribution**.

Sampling Distribution Given simple random samples of size n from a given population with a measured characteristic such as mean, proportion, or standard deviation for each sample, the probability distribution of all the measured characteristics is called a sampling distribution.

Standard Error of the Mean the standard deviation of the distribution of the sample means, or $\frac{\sigma}{\sqrt{n}}$.

Uniform Distribution a continuous random variable (RV) that has equally likely outcomes over the domain, $a < x < b$; often referred as the **Rectangular Distribution** because the graph of the pdf has the form of a rectangle. Notation: $X \sim U(a, b)$. The mean is $\mu = \frac{a+b}{2}$ and the standard deviation is $\sigma = \sqrt{\frac{(b-a)^2}{12}}$. The probability density function is $f(x) = \frac{1}{b-a}$ for $a < x < b$ or $a \leq x \leq b$. The cumulative distribution is $P(X \leq x) = \frac{x-a}{b-a}$.

CHAPTER REVIEW

7.1 The Central Limit Theorem for Sample Means (Averages)

In a population whose distribution may be known or unknown, if the size (n) of samples is sufficiently large, the distribution of the sample means will be approximately normal. The mean of the sample means will equal the population mean. The standard deviation of the distribution of the sample means, called the standard error of the mean, is equal to the population standard deviation divided by the square root of the sample size (n).

7.2 The Central Limit Theorem for Sums

The central limit theorem tells us that for a population with any distribution, the distribution of the sums for the sample means approaches a normal distribution as the sample size increases. In other words, if the sample size is large enough, the distribution of the sums can be approximated by a normal distribution even if the original population is not normally distributed. Additionally, if the original population has a mean of μ_X and a standard deviation of σ_X, the mean of the sums is $n\mu_X$ and the standard deviation is $(\sqrt{n})(\sigma_X)$ where n is the sample size.

7.3 Using the Central Limit Theorem

The central limit theorem can be used to illustrate the law of large numbers. The law of large numbers states that the larger the sample size you take from a population, the closer the sample mean $\bar{x}$ gets to μ.

FORMULA REVIEW

7.1 The Central Limit Theorem for Sample Means (Averages)

The Central Limit Theorem for Sample Means: $\bar{X} \sim N\left(\mu_x, \frac{\sigma x}{\sqrt{n}}\right)$

The Mean $\bar{X}$: μ_x

Central Limit Theorem for Sample Means z-score and standard error of the mean: $z = \frac{\bar{x} - \mu_x}{\left(\frac{\sigma_x}{\sqrt{n}}\right)}$

Standard Error of the Mean (Standard Deviation ($\bar{X}$)): $\frac{\sigma_x}{\sqrt{n}}$

7.2 The Central Limit Theorem for Sums

The Central Limit Theorem for Sums: $\sum X \sim N[(n)(\mu_x),(\sqrt{n})(\sigma_x)]$

Mean for Sums ($\sum X$): $(n)(\mu_x)$

The Central Limit Theorem for Sums z-score and standard deviation for sums: $z\text{for the sample mean} = \frac{\Sigma x - (n)(\mu_X)}{(\sqrt{n})(\sigma_X)}$

Standard deviation for Sums ($\sum X$): $(\sqrt{n})(\sigma_x)$

PRACTICE

7.1 The Central Limit Theorem for Sample Means (Averages)

Use the following information to answer the next six exercises: Yoonie is a personnel manager in a large corporation. Each month she must review 16 of the employees. From past experience, she has found that the reviews take her approximately four hours each to do with a population standard deviation of 1.2 hours. Let X be the random variable representing the time it takes her to complete one review. Assume X is normally distributed. Let $\bar{X}$ be the random variable representing the mean time to complete the 16 reviews. Assume that the 16 reviews represent a random set of reviews.

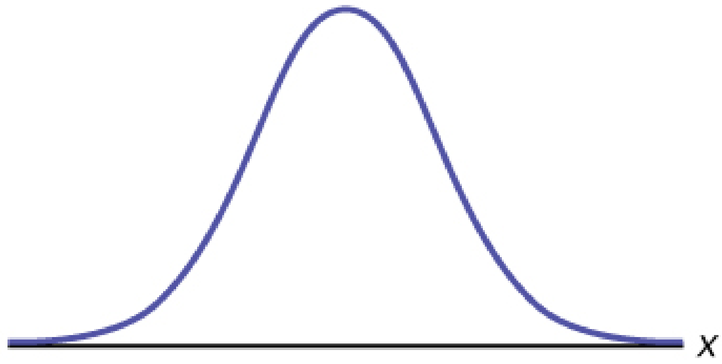
1. What is the mean, standard deviation, and sample size?

2. Complete the distributions.

a. $X \sim$ _____(_____,_____)

b. $\bar{X} \sim$ _____(_____,_____)

3. Find the probability that **one** review will take Yoonie from 3.5 to 4.25 hours. Sketch the graph, labeling and scaling the horizontal axis. Shade the region corresponding to the probability.

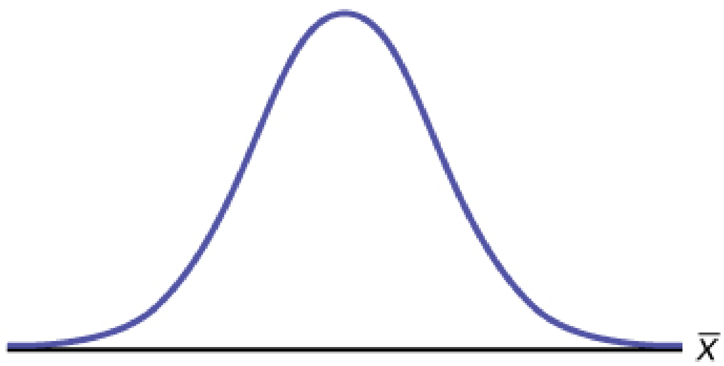


a.

Figure 7.16

b. $P(\_\_\_\_\_\_\_\_ < x < \_\_\_\_\_\_\_\_) = \_\_\_\_\_\_\_$

4. Find the probability that the **mean** of a month’s reviews will take Yoonie from 3.5 to 4.25 hrs. Sketch the graph, labeling and scaling the horizontal axis. Shade the region corresponding to the probability.



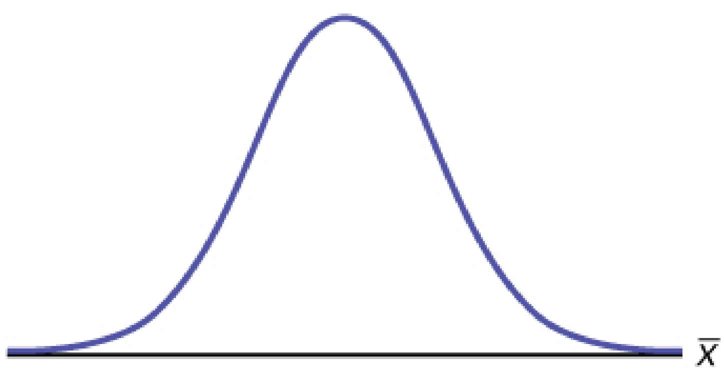
a.

Figure 7.17

b. $P(\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) = \_\_\_\_\_\_\_$

5. What causes the probabilities in **Exercise 7.3** and **Exercise 7.4** to be different?

6. Find the 95th percentile for the mean time to complete one month's reviews. Sketch the graph.



a.

Figure 7.18

b. The 95th Percentile =____________

7.2 The Central Limit Theorem for Sums

Use the following information to answer the next four exercises: An unknown distribution has a mean of 80 and a standard deviation of 12. A sample size of 95 is drawn randomly from the population.

7. Find the probability that the sum of the 95 values is greater than 7,650.

8. Find the probability that the sum of the 95 values is less than 7,400.

9. Find the sum that is two standard deviations above the mean of the sums.

10. Find the sum that is 1.5 standard deviations below the mean of the sums.

Use the following information to answer the next five exercises: The distribution of results from a cholesterol test has a mean of 180 and a standard deviation of 20. A sample size of 40 is drawn randomly.

11. Find the probability that the sum of the 40 values is greater than 7,500.

12. Find the probability that the sum of the 40 values is less than 7,000.

13. Find the sum that is one standard deviation above the mean of the sums.

14. Find the sum that is 1.5 standard deviations below the mean of the sums.

15. Find the percentage of sums between 1.5 standard deviations below the mean of the sums and one standard deviation above the mean of the sums.

Use the following information to answer the next six exercises: A researcher measures the amount of sugar in several cans of the same soda. The mean is 39.01 with a standard deviation of 0.5. The researcher randomly selects a sample of 100.

16. Find the probability that the sum of the 100 values is greater than 3,910.

17. Find the probability that the sum of the 100 values is less than 3,900.

18. Find the probability that the sum of the 100 values falls between the numbers you found in and .

19. Find the sum with a z–score of –2.5.

20. Find the sum with a z–score of 0.5.

21. Find the probability that the sums will fall between the z-scores **–**2 and 1.

Use the following information to answer the next four exercise: An unknown distribution has a mean 12 and a standard deviation of one. A sample size of 25 is taken. Let X = the object of interest.

22. What is the mean of ΣX?

23. What is the standard deviation of ΣX?

24. What is $P(\Sigma x = 290)$?

25. What is $P(\Sigma x > 290)$?

26. True or False: only the sums of normal distributions are also normal distributions.

27. In order for the sums of a distribution to approach a normal distribution, what must be true?

28. What three things must you know about a distribution to find the probability of sums?

29. An unknown distribution has a mean of 25 and a standard deviation of six. Let X = one object from this distribution. What is the sample size if the standard deviation of ΣX is 42?

30. An unknown distribution has a mean of 19 and a standard deviation of 20. Let X = the object of interest. What is the sample size if the mean of ΣX is 15,200?

Use the following information to answer the next three exercises. A market researcher analyzes how many electronics devices customers buy in a single purchase. The distribution has a mean of three with a standard deviation of 0.7. She samples 400 customers.

31. What is the z-score for $\Sigma x = 840$?

32. What is the z-score for $\Sigma x = 1{,}186$?

33. What is $P(\Sigma x < 1{,}186)$?

Use the following information to answer the next three exercises: An unkwon distribution has a mean of 100, a standard deviation of 100, and a sample size of 100. Let X = one object of interest.

34. What is the mean of ΣX?

35. What is the standard deviation of ΣX?

36. What is $P(\Sigma x > 9{,}000)$?

7.3 Using the Central Limit Theorem

Use the following information to answer the next ten exercises: A manufacturer produces 25-pound lifting weights. The lowest actual weight is 24 pounds, and the highest is 26 pounds. Each weight is equally likely so the distribution of weights is uniform. A sample of 100 weights is taken.

37.

a. What is the distribution for the weights of one 25-pound lifting weight? What is the mean and standard deivation?
b. What is the distribution for the mean weight of 100 25-pound lifting weights?
c. Find the probability that the mean actual weight for the 100 weights is less than 24.9.

38. Draw the graph from **Exercise 7.37**

39. Find the probability that the mean actual weight for the 100 weights is greater than 25.2.

40. Draw the graph from **Exercise 7.39**

41. Find the 90th percentile for the mean weight for the 100 weights.

42. Draw the graph from **Exercise 7.41**

43.

a. What is the distribution for the sum of the weights of 100 25-pound lifting weights?
b. Find $P(\Sigma x < 2{,}450)$.

44. Draw the graph from **Exercise 7.43**

45. Find the 90th percentile for the total weight of the 100 weights.

46. Draw the graph from **Exercise 7.45**

Use the following information to answer the next five exercises: The length of time a particular smartphone's battery lasts follows an exponential distribution with a mean of ten months. A sample of 64 of these smartphones is taken.

47.

a. What is the standard deviation?
b. What is the parameter m?

48. What is the distribution for the length of time one battery lasts?

49. What is the distribution for the mean length of time 64 batteries last?

50. What is the distribution for the total length of time 64 batteries last?

51. Find the probability that the sample mean is between seven and 11.

52. Find the 80th percentile for the total length of time 64 batteries last.

53. Find the *IQR* for the mean amount of time 64 batteries last.

54. Find the middle 80% for the total amount of time 64 batteries last.

Use the following information to answer the next eight exercises: A uniform distribution has a minimum of six and a maximum of ten. A sample of 50 is taken.

55. Find $P(\Sigma x > 420)$.

56. Find the 90th percentile for the sums.

57. Find the 15th percentile for the sums.

58. Find the first quartile for the sums.

59. Find the third quartile for the sums.

60. Find the 80th percentile for the sums.

HOMEWORK

7.1 The Central Limit Theorem for Sample Means (Averages)

61. Previously, De Anza statistics students estimated that the amount of change daytime statistics students carry is exponentially distributed with a mean of $0.88. Suppose that we randomly pick 25 daytime statistics students.

a. In words, X = ____________
b. X ~ _____(_____,_____)
c. In words, $\bar{X}$ = ____________
d. $\bar{X}$ ~ ______ (______, ______)
e. Find the probability that an individual had between $0.80 and $1.00. Graph the situation, and shade in the area to be determined.
f. Find the probability that the average of the 25 students was between $0.80 and $1.00. Graph the situation, and shade in the area to be determined.
g. Explain why there is a difference in part e and part f.

62. Suppose that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet. We randomly sample 49 fly balls.

a. If $\bar{X}$ = average distance in feet for 49 fly balls, then $\bar{X}$ ~ _______(_______,_______)
b. What is the probability that the 49 balls traveled an average of less than 240 feet? Sketch the graph. Scale the horizontal axis for $\bar{X}$. Shade the region corresponding to the probability. Find the probability.
c. Find the 80th percentile of the distribution of the average of 49 fly balls.

63. According to the Internal Revenue Service, the average length of time for an individual to complete (keep records for, learn, prepare, copy, assemble, and send) IRS Form 1040 is 10.53 hours (without any attached schedules). The distribution is unknown. Let us assume that the standard deviation is two hours. Suppose we randomly sample 36 taxpayers.

a. In words, X = _____________
b. In words, $\bar{X}$ = _____________
c. $\bar{X}$ ~ _____(_____,_____)
d. Would you be surprised if the 36 taxpayers finished their Form 1040s in an average of more than 12 hours? Explain why or why not in complete sentences.
e. Would you be surprised if one taxpayer finished his or her Form 1040 in more than 12 hours? In a complete sentence, explain why.

64. Suppose that a category of world-class runners are known to run a marathon (26 miles) in an average of 145 minutes with a standard deviation of 14 minutes. Consider 49 of the races. Let $\bar{X}$ the average of the 49 races.

a. $\bar{X}$ ~ _____(_____,_____)
b. Find the probability that the runner will average between 142 and 146 minutes in these 49 marathons.
c. Find the 80th percentile for the average of these 49 marathons.
d. Find the median of the average running times.

65. The length of songs in a collector's iTunes album collection is uniformly distributed from two to 3.5 minutes. Suppose we randomly pick five albums from the collection. There are a total of 43 songs on the five albums.

a. In words, X = _________
b. X ~ _____________
c. In words, $\bar{X}$ = _____________
d. $\bar{X}$ ~ _____(_____,_____)
e. Find the first quartile for the average song length.
f. The IQR(interquartile range) for the average song length is from _______–_______.

66. In 1940 the average size of a U.S. farm was 174 acres. Let's say that the standard deviation was 55 acres. Suppose we randomly survey 38 farmers from 1940.

a. In words, X = _____________
b. In words, $\bar{X}$ = _____________
c. $\bar{X}$ ~ _____(_____,_____)
d. The IQR for $\bar{X}$ is from _______ acres to _______ acres.

67. Determine which of the following are true and which are false. Then, in complete sentences, justify your answers.

a. When the sample size is large, the mean of $\bar{X}$ is approximately equal to the mean of X.

b. When the sample size is large, $\bar{X}$ is approximately normally distributed.

c. When the sample size is large, the standard deviation of $\bar{X}$ is approximately the same as the standard deviation of X.

68. The percent of fat calories that a person in America consumes each day is normally distributed with a mean of about 36 and a standard deviation of about ten. Suppose that 16 individuals are randomly chosen. Let $\bar{X}$ = average percent of fat calories.

a. $\bar{X}$ ~ ______(______, ______)
b. For the group of 16, find the probability that the average percent of fat calories consumed is more than five. Graph the situation and shade in the area to be determined.
c. Find the first quartile for the average percent of fat calories.

69. The distribution of income in some Third World countries is considered wedge shaped (many very poor people, very few middle income people, and even fewer wealthy people). Suppose we pick a country with a wedge shaped distribution. Let the average salary be $2,000 per year with a standard deviation of $8,000. We randomly survey 1,000 residents of that country.

a. In words, X = _____________
b. In words, $\bar{X}$ = _____________
c. $\bar{X}$ ~ _____(_____,_____)
d. How is it possible for the standard deviation to be greater than the average?
e. Why is it more likely that the average of the 1,000 residents will be from $2,000 to $2,100 than from $2,100 to $2,200?

70. Which of the following is NOT TRUE about the distribution for averages?

a. The mean, median, and mode are equal.
b. The area under the curve is one.
c. The curve never touches the x-axis.
d. The curve is skewed to the right.

71. The cost of unleaded gasoline in the Bay Area once followed an unknown distribution with a mean of $4.59 and a standard deviation of $0.10. Sixteen gas stations from the Bay Area are randomly chosen. We are interested in the average cost of gasoline for the 16 gas stations. The distribution to use for the average cost of gasoline for the 16 gas stations is:

a. $\bar{X} \sim N(4.59, 0.10)$
b. $\bar{X} \sim N\left(4.59, \frac{0.10}{\sqrt{16}}\right)$
c. $\bar{X} \sim N\left(4.59, \frac{16}{0.10}\right)$
d. $\bar{X} \sim N\left(4.59, \frac{\sqrt{16}}{0.10}\right)$

7.2 The Central Limit Theorem for Sums

72. Which of the following is NOT TRUE about the theoretical distribution of sums?

a. The mean, median and mode are equal.
b. The area under the curve is one.
c. The curve never touches the x-axis.
d. The curve is skewed to the right.

73. Suppose that the duration of a particular type of criminal trial is known to have a mean of 21 days and a standard deviation of seven days. We randomly sample nine trials.

a. In words, ΣX = ______________
b. ΣX ~ _____(_____,_____)
c. Find the probability that the total length of the nine trials is at least 225 days.
d. Ninety percent of the total of nine of these types of trials will last at least how long?

74. Suppose that the weight of open boxes of cereal in a home with children is uniformly distributed from two to six pounds with a mean of four pounds and standard deviation of 1.1547. We randomly survey 64 homes with children.

a. In words, X = _____________
b. The distribution is _______.
c. In words, ΣX = _______________

d. $\Sigma X \sim$ _____(_____,_____)
e. Find the probability that the total weight of open boxes is less than 250 pounds.
f. Find the 35th percentile for the total weight of open boxes of cereal.

75. Salaries for teachers in a particular elementary school district are normally distributed with a mean of $44,000 and a standard deviation of $6,500. We randomly survey ten teachers from that district.

a. In words, X = ______________
b. $X \sim$ _____(_____,_____)
c. In words, ΣX = ______________
d. $\Sigma X \sim$ _____(_____,_____)
e. Find the probability that the teachers earn a total of over $400,000.
f. Find the 90th percentile for an individual teacher's salary.
g. Find the 90th percentile for the sum of ten teachers' salary.
h. If we surveyed 70 teachers instead of ten, graphically, how would that change the distribution in part d?
i. If each of the 70 teachers received a $3,000 raise, graphically, how would that change the distribution in part b?

7.3 Using the Central Limit Theorem

76. The attention span of a two-year-old is exponentially distributed with a mean of about eight minutes. Suppose we randomly survey 60 two-year-olds.

a. In words, X = _______
b. $X \sim$ _____(_____,_____)
c. In words, $\bar{X}$ = ____________
d. $\bar{X} \sim$ _____(_____,_____)
e. Before doing any calculations, which do you think will be higher? Explain why.
   i. The probability that an individual attention span is less than ten minutes.
   ii. The probability that the average attention span for the 60 children is less than ten minutes?
f. Calculate the probabilities in part e.
g. Explain why the distribution for $\bar{X}$ is not exponential.

77. The closing stock prices of 35 U.S. semiconductor manufacturers are given as follows.

8.625; 30.25; 27.625; 46.75; 32.875; 18.25; 5; 0.125; 2.9375; 6.875; 28.25; 24.25; 21; 1.5; 30.25; 71; 43.5; 49.25; 2.5625; 31; 16.5; 9.5; 18.5; 18; 9; 10.5; 16.625; 1.25; 18; 12.87; 7; 12.875; 2.875; 60.25; 29.25

a. In words, X = ______________
b.
   i. $\bar{x}$ = _____
   ii. s_x = _____
   iii. n = _____
c. Construct a histogram of the distribution of the averages. Start at x = –0.0005. Use bar widths of ten.
d. In words, describe the distribution of stock prices.
e. Randomly average five stock prices together. (Use a random number generator.) Continue averaging five pieces together until you have ten averages. List those ten averages.
f. Use the ten averages from part e to calculate the following.
   i. $\bar{x}$ = _____
   ii. s_x = _____
g. Construct a histogram of the distribution of the averages. Start at x = -0.0005. Use bar widths of ten.
h. Does this histogram look like the graph in part c?
i. In one or two complete sentences, explain why the graphs either look the same or look different?
j. Based upon the theory of the **central limit theorem**, $\bar{X} \sim$ _____(_____,____)

Use the following information to answer the next three exercises: Richard’s Furniture Company delivers furniture from 10 A.M. to 2 P.M. continuously and uniformly. We are interested in how long (in hours) past the 10 A.M. start time that individuals wait for their delivery.

78. $X \sim$ _____(_____,_____)

a. $U(0,4)$
b. $U(10,2)$
c. $E\chi p(2)$

d. $N(2,1)$

79. The average wait time is:
a. one hour.
b. two hours.
c. two and a half hours.
d. four hours.

80. Suppose that it is now past noon on a delivery day. The probability that a person must wait at least one and a half **more** hours is:
a. $\frac{1}{4}$
b. $\frac{1}{2}$
c. $\frac{3}{4}$
d. $\frac{3}{8}$

Use the following information to answer the next two exercises: The time to wait for a particular rural bus is distributed uniformly from zero to 75 minutes. One hundred riders are randomly sampled to learn how long they waited.

81. The 90th percentile sample average wait time (in minutes) for a sample of 100 riders is:
a. 315.0
b. 40.3
c. 38.5
d. 65.2

82. Would you be surprised, based upon numerical calculations, if the sample average wait time (in minutes) for 100 riders was less than 30 minutes?
a. yes
b. no
c. There is not enough information.

Use the following to answer the next two exercises: The cost of unleaded gasoline in the Bay Area once followed an unknown distribution with a mean of $4.59 and a standard deviation of $0.10. Sixteen gas stations from the Bay Area are randomly chosen. We are interested in the average cost of gasoline for the 16 gas stations.

83. What's the approximate probability that the average price for 16 gas stations is over $4.69?
a. almost zero
b. 0.1587
c. 0.0943
d. unknown

84. Find the probability that the average price for 30 gas stations is less than $4.55.
a. 0.6554
b. 0.3446
c. 0.0142
d. 0.9858
e. 0

85. Suppose in a local Kindergarten through 12th grade (K - 12) school district, 53 percent of the population favor a charter school for grades K through five. A simple random sample of 300 is surveyed. Calculate following using the normal approximation to the binomial distribtion.
a. Find the probability that less than 100 favor a charter school for grades K through 5.
b. Find the probability that 170 or more favor a charter school for grades K through 5.
c. Find the probability that no more than 140 favor a charter school for grades K through 5.
d. Find the probability that there are fewer than 130 that favor a charter school for grades K through 5.
e. Find the probability that exactly 150 favor a charter school for grades K through 5.

If you have access to an appropriate calculator or computer software, try calculating these probabilities using the technology.

86. Four friends, Janice, Barbara, Kathy and Roberta, decided to carpool together to get to school. Each day the driver would be chosen by randomly selecting one of the four names. They carpool to school for 96 days. Use the normal approximation to the binomial to calculate the following probabilities. Round the standard deviation to four decimal places.

a. Find the probability that Janice is the driver at most 20 days.

b. Find the probability that Roberta is the driver more than 16 days.
c. Find the probability that Barbara drives exactly 24 of those 96 days.

87. $X \sim N(60, 9)$. Suppose that you form random samples of 25 from this distribution. Let $\bar{X}$ be the random variable of averages. Let ΣX be the random variable of sums. For parts c through f, sketch the graph, shade the region, label and scale the horizontal axis for $\bar{X}$, and find the probability.

a. Sketch the distributions of X and $\bar{X}$ on the same graph.
b. $\bar{X} \sim$ _____(_____,_____)
c. $P(\bar{x} < 60)$ = _____
d. Find the 30th percentile for the mean.
e. $P(56 < \bar{x} < 62)$ = _____
f. $P(18 < \bar{x} < 58)$ = _____
g. $\Sigma x \sim$ _____(_____,_____)
h. Find the minimum value for the upper quartile for the sum.
i. $P(1{,}400 < \Sigma x < 1{,}550)$ = _____

88. Suppose that the length of research papers is uniformly distributed from ten to 25 pages. We survey a class in which 55 research papers were turned in to a professor. The 55 research papers are considered a random collection of all papers. We are interested in the average length of the research papers.

a. In words, X = _____________
b. $X \sim$ _____(_____,_____)
c. μ_X = _____
d. σ_X = _____
e. In words, $\bar{X}$ = ______________
f. $\bar{X} \sim$ _____(_____,_____)
g. In words, ΣX = _____________
h. $\Sigma X \sim$ _____(_____,_____)
i. Without doing any calculations, do you think that it's likely that the professor will need to read a total of more than 1,050 pages? Why?
j. Calculate the probability that the professor will need to read a total of more than 1,050 pages.
k. Why is it so unlikely that the average length of the papers will be less than 12 pages?

89. Salaries for teachers in a particular elementary school district are normally distributed with a mean of \$44,000 and a standard deviation of \$6,500. We randomly survey ten teachers from that district.

a. Find the 90th percentile for an individual teacher's salary.
b. Find the 90th percentile for the average teacher's salary.

90. The average length of a maternity stay in a U.S. hospital is said to be 2.4 days with a standard deviation of 0.9 days. We randomly survey 80 women who recently bore children in a U.S. hospital.

a. In words, X = _____________
b. In words, $\bar{X}$ = ___________________
c. $\bar{X} \sim$ _____(_____,_____)
d. In words, ΣX = _______________
e. $\Sigma X \sim$ _____(_____,_____)
f. Is it likely that an individual stayed more than five days in the hospital? Why or why not?
g. Is it likely that the average stay for the 80 women was more than five days? Why or why not?
h. Which is more likely:
    i. An individual stayed more than five days.
    ii. the average stay of 80 women was more than five days.
i. If we were to sum up the women's stays, is it likely that, collectively they spent more than a year in the hospital? Why or why not?

For each problem, wherever possible, provide graphs and use the calculator.

91. NeverReady batteries has engineered a newer, longer lasting AAA battery. The company claims this battery has an average life span of 17 hours with a standard deviation of 0.8 hours. Your statistics class questions this claim. As a class, you randomly select 30 batteries and find that the sample mean life span is 16.7 hours. If the process is working properly,

what is the probability of getting a random sample of 30 batteries in which the sample mean lifetime is 16.7 hours or less? Is the company's claim reasonable?

92. Men have an average weight of 172 pounds with a standard deviation of 29 pounds.

a. Find the probability that 20 randomly selected men will have a sum weight greater than 3600 lbs.
b. If 20 men have a sum weight greater than 3500 lbs, then their total weight exceeds the safety limits for water taxis. Based on (a), is this a safety concern? Explain.

93. M&M candies large candy bags have a claimed net weight of 396.9 g. The standard deviation for the weight of the individual candies is 0.017 g. The following table is from a stats experiment conducted by a statistics class.

Red	Orange	Yellow	Brown	Blue	Green
0.751	0.735	0.883	0.696	0.881	0.925
0.841	0.895	0.769	0.876	0.863	0.914
0.856	0.865	0.859	0.855	0.775	0.881
0.799	0.864	0.784	0.806	0.854	0.865
0.966	0.852	0.824	0.840	0.810	0.865
0.859	0.866	0.858	0.868	0.858	1.015
0.857	0.859	0.848	0.859	0.818	0.876
0.942	0.838	0.851	0.982	0.868	0.809
0.873	0.863			0.803	0.865
0.809	0.888			0.932	0.848
0.890	0.925			0.842	0.940
0.878	0.793			0.832	0.833
0.905	0.977			0.807	0.845
	0.850			0.841	0.852
	0.830			0.932	0.778
	0.856			0.833	0.814
	0.842			0.881	0.791
	0.778			0.818	0.810
	0.786			0.864	0.881
	0.853			0.825	
	0.864			0.855	
	0.873			0.942	
	0.880			0.825	
	0.882			0.869	
	0.931			0.912	
				0.887	

Table 7.7

The bag contained 465 candies and he listed weights in the table came from randomly selected candies. Count the weights.

a. Find the mean sample weight and the standard deviation of the sample weights of candies in the table.
b. Find the sum of the sample weights in the table and the standard deviation of the sum the of the weights.
c. If 465 M&Ms are randomly selected, find the probability that their weights sum to at least 396.9.
d. Is the Mars Company's M&M labeling accurate?

94. The Screw Right Company claims their $\frac{3}{4}$ inch screws are within ±0.23 of the claimed mean diameter of 0.750 inches with a standard deviation of 0.115 inches. The following data were recorded.

0.757	0.723	0.754	0.737	0.757	0.741	0.722	0.741	0.743	0.742
0.740	0.758	0.724	0.739	0.736	0.735	0.760	0.750	0.759	0.754
0.744	0.758	0.765	0.756	0.738	0.742	0.758	0.757	0.724	0.757
0.744	0.738	0.763	0.756	0.760	0.768	0.761	0.742	0.734	0.754
0.758	0.735	0.740	0.743	0.737	0.737	0.725	0.761	0.758	0.756

Table 7.8

The screws were randomly selected from the local home repair store.

a. Find the mean diameter and standard deviation for the sample
b. Find the probability that 50 randomly selected screws will be within the stated tolerance levels. Is the company's diameter claim plausible?

95. Your company has a contract to perform preventive maintenance on thousands of air-conditioners in a large city. Based on service records from previous years, the time that a technician spends servicing a unit averages one hour with a standard deviation of one hour. In the coming week, your company will service a simple random sample of 70 units in the city. You plan to budget an average of 1.1 hours per technician to complete the work. Will this be enough time?

96. A typical adult has an average IQ score of 105 with a standard deviation of 20. If 20 randomly selected adults are given an IQ tesst, what is the probability that the sample mean scores will be between 85 and 125 points?

97. Certain coins have an average weight of 5.201 grams with a standard deviation of 0.065 g. If a vending machine is designed to accept coins whose weights range from 5.111 g to 5.291 g, what is the expected number of rejected coins when 280 randomly selected coins are inserted into the machine?

REFERENCES

7.1 The Central Limit Theorem for Sample Means (Averages)

Baran, Daya. "20 Percent of Americans Have Never Used Email."WebGuild, 2010. Available online at http://www.webguild.org/20080519/20-percent-of-americans-have-never-used-email (accessed May 17, 2013).

Data from The Flurry Blog, 2013. Available online at http://blog.flurry.com (accessed May 17, 2013).

Data from the United States Department of Agriculture.

7.2 The Central Limit Theorem for Sums

Farago, Peter. "The Truth About Cats and Dogs: Smartphone vs Tablet Usage Differences." The Flurry Blog, 2013. Posted October 29, 2012. Available online at http://blog.flurry.com (accessed May 17, 2013).

7.3 Using the Central Limit Theorem

Data from the Wall Street Journal.

"National Health and Nutrition Examination Survey." Center for Disease Control and Prevention. Available online at http://www.cdc.gov/nchs/nhanes.htm (accessed May 17, 2013).

SOLUTIONS

1 mean = 4 hours; standard deviation = 1.2 hours; sample size = 16

3 a. Check student's solution.
b. 3.5, 4.25, 0.2441

5 The fact that the two distributions are different accounts for the different probabilities.

7 0.3345

9 7,833.92

11 0.0089

13 7,326.49

15 77.45%

17 0.4207

19 3,888.5

21 0.8186

23 5

25 0.9772

27 The sample size, *n*, gets larger.

29 49

31 26.00

33 0.1587

35 1,000

37

a. $U(24, 26)$, 25, 0.5774
b. $N(25, 0.0577)$
c. 0.0416

39 0.0003

41 25.07

43

a. $N(2,500, 5.7735)$
b. 0

45 2,507.40

47

a. 10
b. $\frac{1}{10}$

49 $N\left(10, \frac{10}{8}\right)$

51 0.7799

53 1.69

55 0.0072

57 391.54

59 405.51

61

a. X = amount of change students carry
b. $X \sim E(0.88, 0.88)$
c. $\bar{X}$ = average amount of change carried by a sample of 25 sstudents.
d. $\bar{X} \sim N(0.88, 0.176)$
e. 0.0819
f. 0.1882
g. The distributions are different. Part a is exponential and part b is normal.

63

a. length of time for an individual to complete IRS form 1040, in hours.
b. mean length of time for a sample of 36 taxpayers to complete IRS form 1040, in hours.
c. $N\left(10.53, \frac{1}{3}\right)$
d. Yes. I would be surprised, because the probability is almost 0.
e. No. I would not be totally surprised because the probability is 0.2312

65

a. the length of a song, in minutes, in the collection
b. $U(2, 3.5)$
c. the average length, in minutes, of the songs from a sample of five albums from the collection
d. $N(2.75, 0.0220)$
e. 2.74 minutes
f. 0.03 minutes

67

a. True. The mean of a sampling distribution of the means is approximately the mean of the data distribution.
b. True. According to the Central Limit Theorem, the larger the sample, the closer the sampling distribution of the means becomes normal.
c. The standard deviation of the sampling distribution of the means will decrease making it approximately the same as the standard deviation of X as the sample size increases.

69

a. X = the yearly income of someone in a third world country
b. the average salary from samples of 1,000 residents of a third world country
c. $\bar{X} \sim N\left(2000, \frac{8000}{\sqrt{1000}}\right)$
d. Very wide differences in data values can have averages smaller than standard deviations.
e. The distribution of the sample mean will have higher probabilities closer to the population mean.

   $P(2000 < \bar{X} < 2100) = 0.1537$

   $P(2100 < \bar{X} < 2200) = 0.1317$

71 b

73

a. the total length of time for nine criminal trials

b. $N(189, 21)$

c. 0.0432

d. 162.09; ninety percent of the total nine trials of this type will last 162 days or more.

75

a. X = the salary of one elementary school teacher in the district

b. $X \sim N(44{,}000, 6{,}500)$

c. ΣX ~ sum of the salaries of ten elementary school teachers in the sample

d. $\Sigma X \sim N(44000, 20554.80)$

e. 0.9742

f. $52,330.09

g. 466,342.04

h. Sampling 70 teachers instead of ten would cause the distribution to be more spread out. It would be a more symmetrical normal curve.

i. If every teacher received a $3,000 raise, the distribution of X would shift to the right by $3,000. In other words, it would have a mean of $47,000.

77

a. X = the closing stock prices for U.S. semiconductor manufacturers

b. i. $20.71; ii. $17.31; iii. 35

d. Exponential distribution, $X \sim Exp\left(\frac{1}{20.71}\right)$

e. Answers will vary.

f. i. $20.71; ii. $11.14

g. Answers will vary.

h. Answers will vary.

i. Answers will vary.

j. $N\left(20.71, \frac{17.31}{\sqrt{5}}\right)$

79 b

81 b

83 a

85

a. 0

b. 0.1123

c. 0.0162

d. 0.0003

e. 0.0268

87

a. Check student's solution.

b. $\bar{X} \sim N\left(60, \frac{9}{\sqrt{25}}\right)$

c. 0.5000
d. 59.06
e. 0.8536
f. 0.1333
g. $N(1500, 45)$
h. 1530.35
i. 0.6877

89

a. $52,330
b. $46,634

91

- We have $\mu = 17$, $\sigma = 0.8$, $\bar{x} = 16.7$, and $n = 30$. To calculate the probability, we use `normalcdf`(lower, upper, μ, $\frac{\sigma}{\sqrt{n}}$) = `normalcdf` $\left(E-99,16.7,17,\frac{0.8}{\sqrt{30}}\right) = 0.0200$.
- If the process is working properly, then the probability that a sample of 30 batteries would have at most 16.7 lifetime hours is only 2%. Therefore, the class was justified to question the claim.

93

a. For the sample, we have $n = 100$, $\bar{x} = 0.862$, $s = 0.05$
b. $\Sigma\bar{x} = 85.65$, $\Sigma s = 5.18$
c. `normalcdf`$(396.9,E99,(465)(0.8565),(0.05)(\sqrt{465})) \approx 1$
d. Since the probability of a sample of size 465 having at least a mean sum of 396.9 is appproximately 1, we can conclude that Mars is correctly labeling their M&M packages.

95 Use `normalcdf` $\left(E-99,1.1,1,\frac{1}{\sqrt{70}}\right) = 0.7986$. This means that there is an 80% chance that the service time will be less than 1.1 hours. It could be wise to schedule more time since there is an associated 20% chance that the maintenance time will be greater than 1.1 hours.

97 Since we have `normalcdf` $\left(5.111,5.291,5.201,\frac{0.065}{\sqrt{280}}\right) \approx 1$, we can conclude that practically all the coins are within the limits, therefore, there should be no rejected coins out of a well selected sample of size 280.